Applied Multivariate Analysis

Homework for Chapter 2

Shuowen Wei

1.

> M

col1 col2

[1,] 0.9 4.9

[2,] 1.0 3.0

[3,] 0.8 1.0

[4,] 2.0 2.9

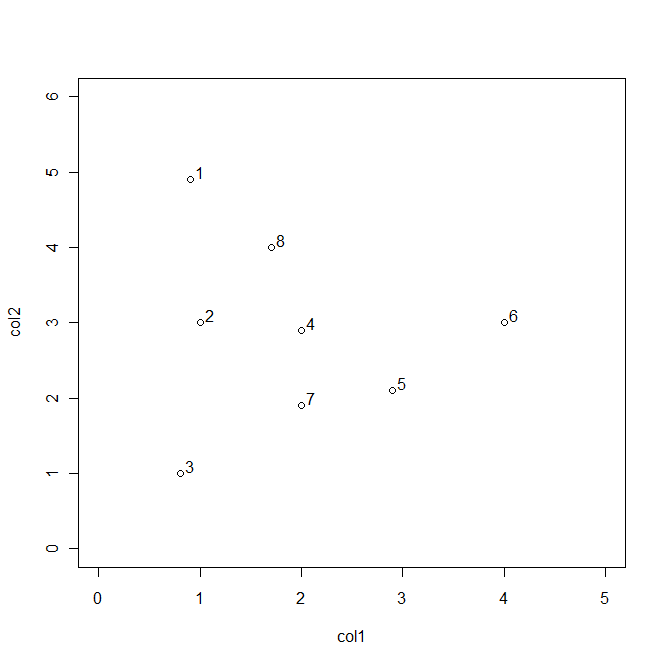
[5,] 2.9 2.1

[6,] 4.0 3.0

[7,] 2.0 1.9

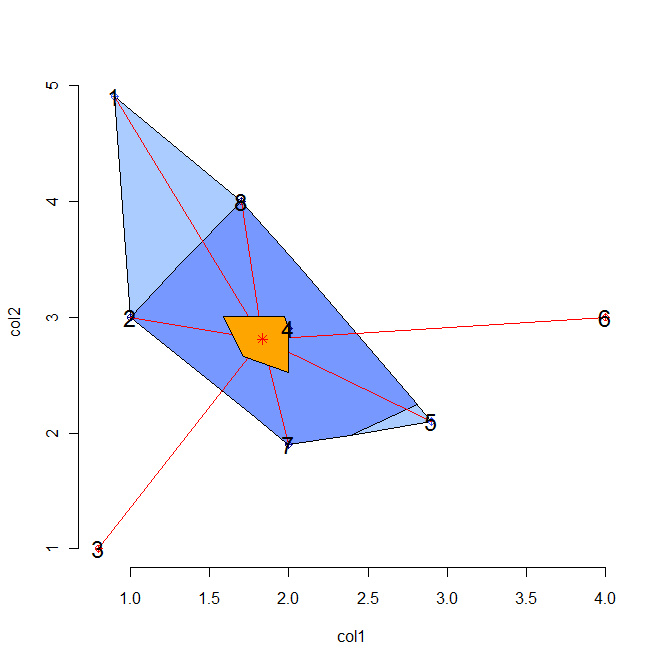
[8,] 1.7 4.0

> plot(M,xlim=c(0,5),ylim=c(0,6));text(M[,1]+0.1,M[,2]+0.1,1:8,cex=1)



2.

> bagplot(M[,1:2],cex=1,pch=10);text(M[,1],M[,2],1:8,cex=1.5)



Thus the bag of the data set if (2,4,7,8).

3.

> chull(M)

[1] 5 3 1 6

4.

By the result of 3, before removing the convex hull, the correlation between the variables is:

> cor(M[,1],M[,2])

[1] -0.09604193

before removing the convex hull,

> mm<-M[-c(1,3,5,6),]

> cor(mm[,1],mm[,2])

[1] -0.3005184

5.

For the point (1,3), we have:

> chi1<-matrix(0,7,2)

> fix(chi1)

> chi1

col1 col2

[1,] 0 0

[2,] 1 0

[3,] 1 0

[4,] 1 0

[5,] 1 1

[6,] 1 1

[7,] 0 1

> cor(chi1[,1],chi1[,2])

[1] -0.09128709

> lambuda1<-4\*(-1)\*max((5/7-1/2)^2,(3/7-1/2)^2)

> lambuda1

[1] -0.1836735

For the point (2,1.9), we have:

> chi2<-matrix(0,7,2)

> fix(chi2)

> chi2

col1 col2

[1,] 0 0

[2,] 1 1

[3,] 1 1

[4,] 1 1

[5,] 0 1

[6,] 0 1

[7,] 0 1

> cor(chi2[,1],chi2[,2])

[1] 0.3535534

> lambuda1<-4\*(-1)\*max((3/7-1/2)^2,(6/7-1/2)^2)

> lambuda1

[1] -0.5102041

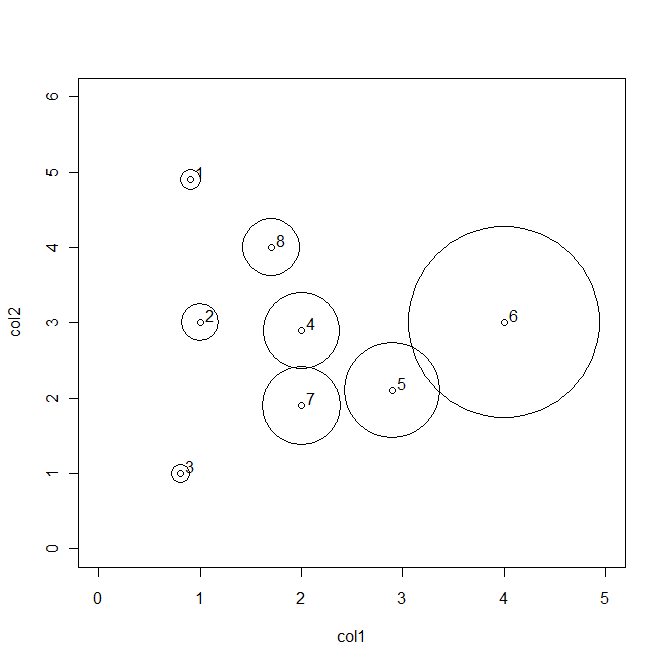
5.

> z

[1] 1.1 2.0 1.0 4.0 5.0 10.0 4.1 3.0

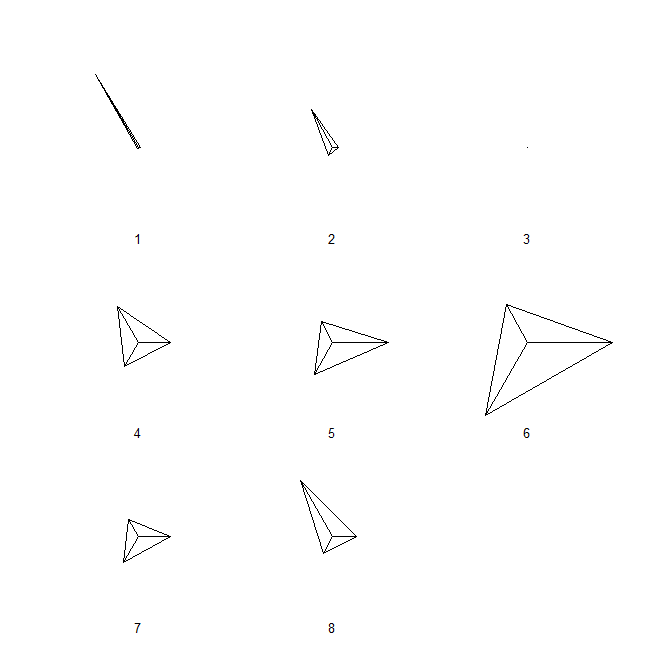
> plot(M,xlim=c(0,5),ylim=c(0,6));text(M[,1]+0.1,M[,2]+0.1,1:8,cex=1)

> symbols(M,circles=z,add=TRUE)



6.

> stars(cbind(M,z),labels=1:8)



8.

> ker

function (x,h)

{

xgrid<-seq(min(x)-3\*h,max(x)+3\*h,0.05)

bumps<-sapply(x,function(y)gauss((xgrid-y)/h))

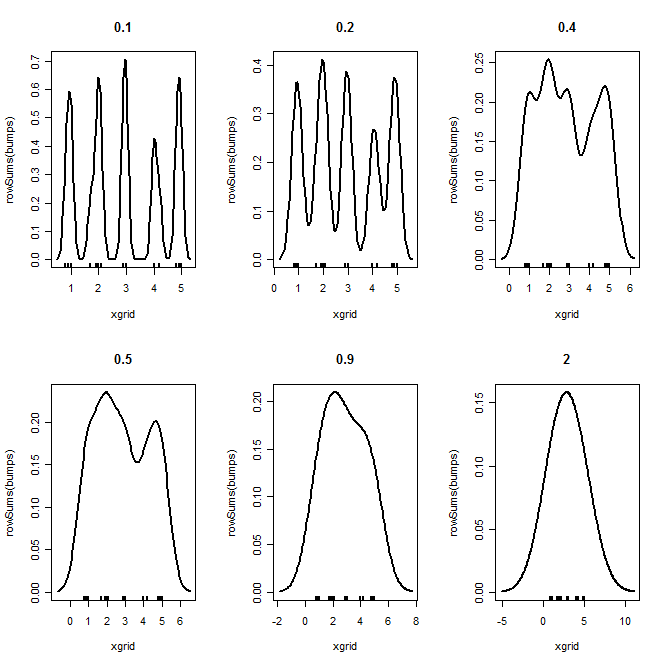
rownames(bumps)<-xgrid;colnames(bumps)<-x

bumps<-bumps/(h\*length(x))

plot(xgrid,rowSums(bumps),ylim=c(0,max(rowSums(bumps))),type="l",lwd=2,main=h)

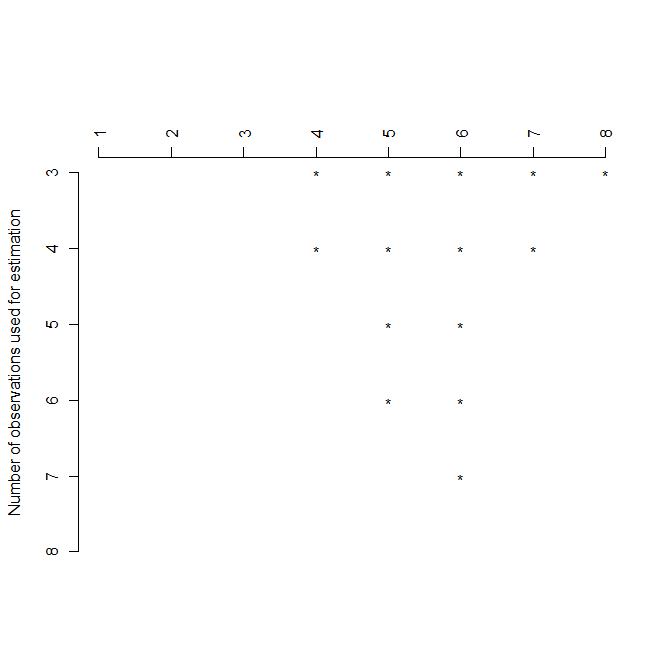
rug(x,lwd=2)

}

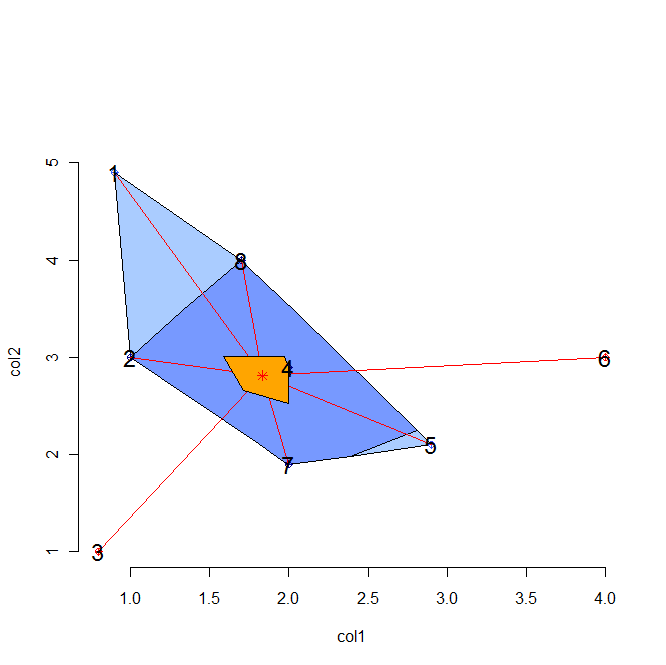


9.

> stalac(M)



> bagplot(M[,1:2],cex=1,pch=10);text(M[,1],M[,2],1:8,cex=1.5)



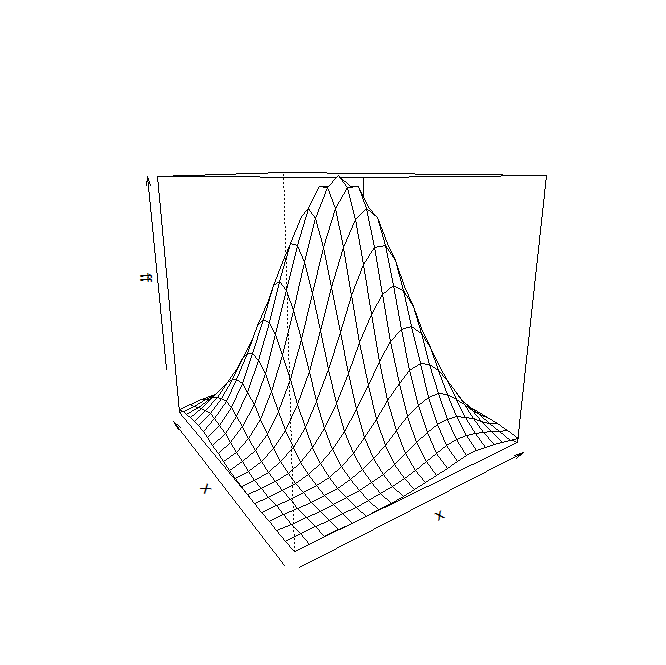
The main issue here is that the way command stalac works. It will first compute the first (q+1) (q is the number of variables, here q=2)variables’ means and covariances and then plug the rest one by one, which means the first (q+1) will have a big affection to the rest. In this case, we can actually see that point 1, 2, 3, point 1 and point 3 are very far away from the others, the multivariate mean and covariance estimated from them exist a big error from the true value.

10.

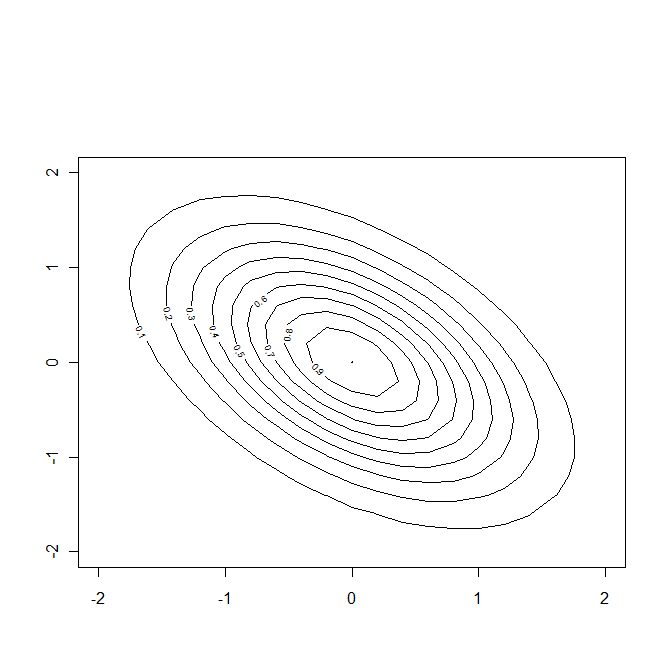
> f<-function(x,y)exp(-(x^2+x\*y+y^2))

>ff<-sapply(x,function(y)f(y,x))

> persp(x,x,ff,theta=-35)

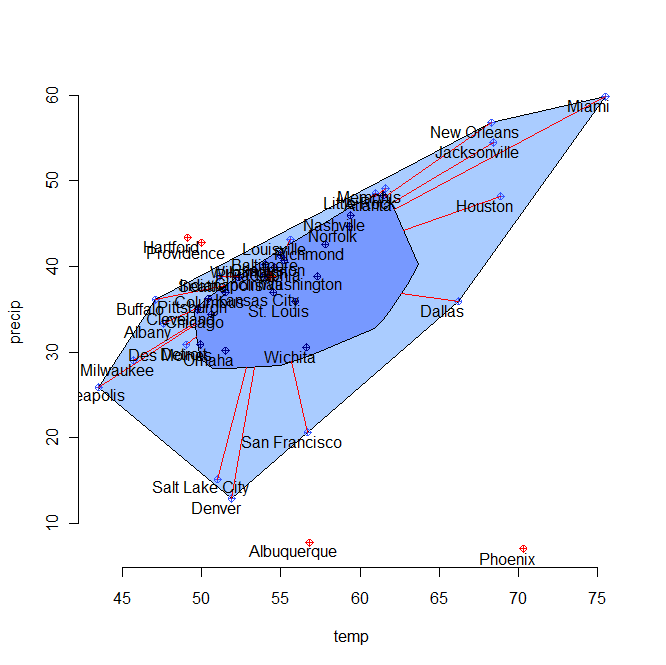


> contour(x,x,ff)



11.

The pair of variables in the USa data set are “temp” and “”



> a<-chull(USa)

> rownames(USa[a,])

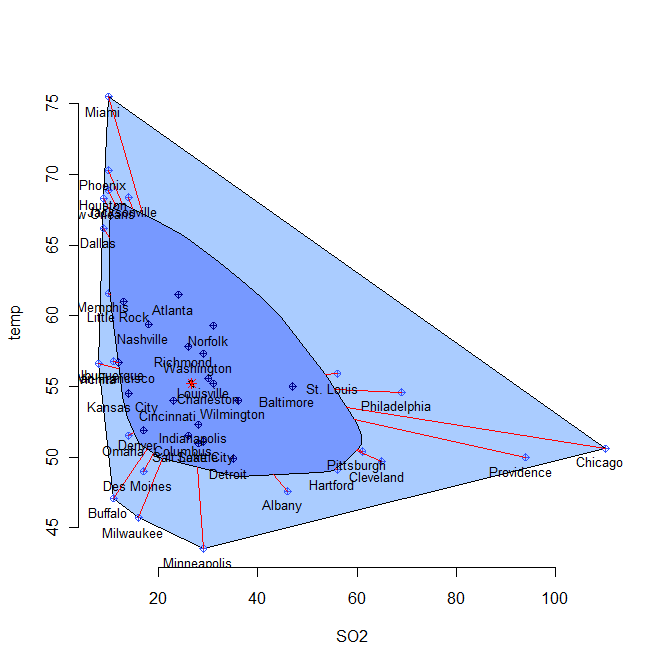
[1] "Minneapolis" "Milwaukee" "Buffalo" "Wichita" "New Orleans" "Miami"

[7] "Chicago"

Those cities above are on the convex hull.

11.

> bagplot(USa[,c(1,2)],cex=1,pch=10);text(USa[,c(1,2)]-1,rownames(USa),cex=0.8)



The pair of variables in the USa data set for which where are no outliers are “SO2” and “temp”.